# Results from the 2020 Census Group Quarters Count Imputation

## *A New Design for the 21st Century*

*Issued May 1, 2023*
*Version 1.0*
*Prepared by Timothy Kennel, Patrick J. Cantwell, Andrew Keller, and Juli Zamora*

*The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. (CBDRB-FY23-078)*

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

United States™ Census 2020

# Group Quarters Imputation Methodology

## Table of Contents

## Table of Tables

# Executive Summary

After the 2020 Census data collection for group quarters had closed, the responses were reviewed. During this review, we discovered that some occupied group quarters lacked a population count. Thus, the Group Quarters Count Imputation Team was formed to develop procedures to impute a population size for these occupied group quarters without a population count. This memo provides results of the GQ count imputation process.

The population size for 5,500 GQs was imputed. This is about 2 percent of the GQ universe, although the unresolved rate varied by GQ type and state. Imputations resulted in adding 169,000 people to the 5,500 GQs.

Although there are some exceptions, the imputed population size generally relied on responses from the Group Quarters Advance Contact operation or listings from current household surveys at the Census Bureau such as the American Community Survey and Current Population Survey.

# Introduction

This document was written to inform the public of results of the 2020 Census Group Quarters (GQs) count imputation process. The 2020 Census GQ count imputation process was used to provide a population size for occupied GQs that lacked a reported population size.

In the 2010 Census, all occupied GQs had a reported population size; thus, no imputation of population size was needed for GQs in the 2010 Census. Additionally, no plans to impute population size for GQs were made during the decade leading up to the 2020 Census. After the 2020 Census data collection had ended and the GQ response was being reviewed, occupied GQs without a population count were found. Thus, the GQ Count Imputation Team was formed in December 2020 and charged with building a statistically sound procedure for imputing population counts for occupied GQs without a population count.

This memo reports results of the GQ count imputation procedure. United States Census Bureau (2022b) provides an overview of the GQ count imputation procedure.

## Group Quarters Count Imputation Description

The GQ count imputation procedure was conducted in February 2021. This activity was conducted during the 2020 Census Response Processing Operation (United States Census Bureau 2022a).

The GQ count imputation procedure was performed on a GQ universe constructed from an intermediate file used during the Decennial Response File 2 creation. The results of the GQ count imputation fed into the Census Unedited File. However, because concurrent and subsequent processing to remove duplicate enumerations in the census and correct other errors altered some of the population counts, the *results in this memo do not represent the final GQ universe that was included in the apportionment counts.* Thus, rates and percentages in this memo are based on denominators that may differ from published numbers of group quarters and people in group quarters. Counts derived from multiplying the rates and percentages in this report by final Census counts should not be attributed to the Census Bureau. Nevertheless, these results are very close to the final counts because the concurrent and subsequent processing had a very small impact on the GQ universe.

There were 5,500 occupied group quarters (GQs) in the GQ universe without a population count.[1] Errors in individual GQ counts, especially for larger GQs, could adversely impact the quality of the census. Thus, for large occupied GQs, a count of 0 is especially problematic. Large

[1] The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. (DRB Approval Number CBDRB-FY23-078). All counts and ratios of GQs have been rounded. Random noise has been added to counts and rates of people in GQs and then rounded.

GQs are often the only living quarters in their tabulation census block. Consequently, information suggesting that an occupied GQ had no population would be evident in the PL 94-171 redistricting data summary file.

This memo walks through the steps used to impute population counts and provides results from each step. The "Imputation Methods" section of this report introduces the imputation methods. The "Determining the Imputation Parameters" section covers the creation of imputation parameters. It describes how the imputation base was created, how data were edited, how ratios were calculated, and how unresolved GQs were edited. The "Results" section provides descriptive statistics summarizing the results of the imputation.

## Operational Changes Resulting from COVID-19

To some extent, the GQ count imputation was developed in response to the COVID-19 pandemic. The closure of many group quarters, such as college student housing, prior to the 2020 Census reference date of April 1, 2020, certainly contributed to some occupied group quarters having a missing population count. In some cases, GQ contacts could not be reached during the primary GQ enumeration because of the pandemic. In other cases, there may have been some confusion about the reference date for the GQ population count, even though GQ contacts were instructed to provide counts prior to closing because of COVID-19.

Although COVID-19 certainly contributed to the presence of some occupied GQs without population counts, it is unclear how many of the occupied GQs did not have population counts because of COVID-19. There certainly could be other reasons why occupied GQs did not have a population count.

## Schedule

The Group Quarters Count Imputation Team was formed in December 2020 to develop and implement a method to impute a positive GQ population count for occupied GQs without a reported Census Day population. To meet the deadline for delivering 2020 Census counts for apportionment, the team was given several weeks to develop a new imputation methodology. Because other postprocessing techniques were being developed to address anomalies in the GQ response data at the same time as the GQ count imputation methods, the universe (including the definitions of the imputation base and unresolved GQs) changed during the weeks when the imputation methodology was being developed and implemented. For more information on census data reviews and anomalies, see Thieme (2021). Because occupied GQs without reported population counts and other anomalies were not anticipated, there was little time to research and develop new methods and software to address the anomalies.

# Background

In the 2020 Decennial Census, the Census Bureau planned two group quarters (GQ) operations, the Group Quarters Advance Contact (GQAC) and the Group Quarters Enumeration (GQE) (refer to United States Census Bureau [2021] for more information on the GQAC and GQE operations). In early 2020, prior to Census Day, the Bureau conducted the GQAC. In this operation, we contacted someone at each GQ location by phone or personal visit to make arrangements for collecting census data around the reference time of the census, Census Day, April 1, 2020. During the GQE, we attempted to collect census data from residents of those GQs based on the enumeration selected by the GQ administrator during GQAC. The GQAC was completed on schedule, ahead of the main onset of the COVID-19 pandemic.

The electronic response (or eResponse) option of the GQE operation was made available on schedule in April 2020 for GQs that selected that option. The Census Bureau also contacted GQ adminstrators who selected in-field options to offer them eResponse, because the Census Bureau was forced to postpone the in-field portion of the operation to midsummer because of the pandemic. At that time, this operation experienced complications because (1) it was difficult to reach many of our contacts at the GQs, and (2) many of the people normally living at GQs around Census Day had moved out because of the pandemic or for other reasons. *It should be noted that, during the GQE, the Census Bureau made it clear that the appropriate reference day for the collection was still April 1, 2020* (with minor exceptions for a few specified types of GQs, such as those in Service-Based Enumeration). *For GQs that closed before April 1 because of the pandemic, the Census Bureau requested the preclosure data*. Further risking the completion and quality of the data collection, in order to compress the schedule for postprocessing the census data, the Census Bureau modified the Late GQE operation, whereby our partners would have assisted in a more thorough review of the GQ response data.

During the planned postprocessing reviews of the data collected from GQs, the Census Bureau observed indications that many GQs were occupied on Census Day but had no population count. These cases represented a variety of circumstances, including reports that the GQ was occupied on Census Day with no population records; that the GQ was vacant at the time of the GQE (midsummer), but occupied on Census Day; and some response refusals with an indication that the GQ was occupied. If not corrected, the issues could lead to an undercount of the people in these GQs; if not properly coordinated with procedures to remove duplicates, they could contribute to an overcount. Beyond these situations, we saw other potential responses in which there appeared to be issues—some based on nonresponse, others based on a report that did not follow the instructions for responding. These issues were not isolated to one area of the country or one type of GQ. Rather, they were found generally in every state and in all or most types of GQs.

To address situations involving the responses from GQs, the Census Bureau first conducted a telephone operation—trying to contact many of the GQs that provided us problematic or no data. Simultaneously, we assembled a team to correct responses when possible, and to apply a new count imputation procedure for many cases in which more specific information was not available. Many subject matter experts on the team reviewed and identified responses with potential errors. For example, every student housing facility is considered a separate GQ and some universities gave us an administrative list of all their students living in university student housing facilities, but inserted all of them into one of the student housing facilities d instead of the student housing where they lived. Although the total number of students living in student housing facilities may be accurate, it is important for us to assign people in the census to the correct block. A few universities provided lists of students assigned to the appropriate GQ, but then included another GQ representing the entire collection of students, essentially doubling the true count. Other problems were detected and corrected where possible.

For a number of GQs, we didn't have response data or subject matter knowledge to determine precisely how many people should be assigned to an occupied GQ. For these cases, we devised business rules to determine for which GQs it was appropriate to apply a GQ count imputation (GQCI) procedure. We then developed, tested, and conducted a GQCI operation on the pertinent cases. Unlike the count imputation procedure for housing unit addresses, which was conducted in prior censuses and used a nearest-neighbor approach in the 2020 Census, the GQCI procedure did not use a nearest-neighbor approach. More like imputation procedures used in economic surveys and censuses, it generally used information available on the GQ under consideration and on the GQs in the same GQ type to impute a count.

## Comprehensive Approach

Where possible, we have taken appropriate steps to correct many of the concerns encountered with GQs. However, these situations were challenging for several reasons. Although one can sometimes determine the likely cause of the problem when the GQ size is large and the subject matter is clear or understandable, for most cases, there were several competing scenarios that may have created the problem—each demanding a different solution. Further, because of the large number of GQs in the country, direct subject matter review could not cover all situations that appeared questionable. Statistical and automated queries of the processed data assisted our research, but they could not ensure that we captured every issue.

To address the problem where we had a GQE population count of 0 and also evidence that the GQ was occupied, we implemented an approach with several components. First, we decreased the number of these cases by conducting a series of processing operations based on data we already had or obtained later, and direct subject matter reviews. These processes resolved many issues. For the remainder of GQs, we applied count imputation.

In early December 2020, the Census Bureau called GQs that had not responded to collect additional information on their population counts around the time of the census. This data collection operation focused on GQs of the following types: state prisons, local jails and other municipal confinement facilities, nursing facilities, college and university student housing, and military quarters. Processing staff then reviewed the newly collected data and placed codes on a file indicating which GQs did not need count imputation because the GQ had provided a good positive count or a valid response of vacant on Census Day during this follow-up operation.

Other investigative research, including a review of notes from the GQAC operation, allowed us to determine some GQs with a valid count of 0. Additional research into GQs classified as correctional facilities provided useful supplemental information. For example, the U.S. Immigration and Customs Enforcement houses some of its detainees at local jails or other facilities and maintains separate records from the local law enforcement. This explained why some of these GQs might legitimately have had a population count of 0 or something small, even though it might appear to be occupied.

Review of many cases found that GQs themselves were often duplicated. That is, GQs with the same name had different addresses, but appeared to be the same GQ. In some cases, the two versions had identical population counts greater than 0; at other times, one of the two had a population count of 0. For these cases, we applied procedures that removed likely duplicate GQs to avoid duplicating the count or imputing a positive count that would have essentially added duplicate records.

Extensive investigation using public data available from the U.S. Department of Education demonstrated a series of problems with reported population counts from colleges and universities in the GQ Enumeration. In some cases, universities provided a plausible count of their student population, but reported that count in a single student housing facility, rather than the student housing facility where the student lived.. Aggregated responses like these can create errors for purposes of redistricting because the people in each GQ should be placed in that GQ and assigned to the correct block. In other cases, the university respondent provided duplicate counts or duplicate person records that inflated the counts of their GQs. If not corrected, this practice could lead to incorrect apportionment counts—an overcount of people in these facilities and geographic areas.

It should be noted that many of these problems occurred in other types of GQs, not just for university housing. After identifying and researching such issues, the team developed procedures that (1) removed person records that were duplicated across different GQs, and (2) reallocated these total counts across the many GQs within the individual GQ facility, such as a university. This helped to spread the total population among the correct GQs—and the correct locations—and prevented the entire count from being erroneously assigned to one block.

In addition, review of the many types of GQs demonstrated that circumstances vary greatly from one type to another. Some GQ facilities tend to be seasonal and may well be truly vacant

for much of the year, including around Census Day. For example, some military quarters tend to open for training according to weather and other factors and were vacant on Census Day. Among GQs visited in the Service-Based Enumeration (SBE) (shelters, soup kitchens, mobile food vans, others), some facilities open and close quickly or according to the season or time of the week. (Note: The Census Bureau's SBE operation was conducted on September 23, 2020, with that day as the reference day—not April 1, 2020.) For these and other reasons, we determined that it would be inappropriate to include some types of GQs in the count imputation process.

As previously mentioned, the Census Bureau conducted an intensive review of the GQs and the distributions of their reported population counts. This allowed the Bureau to develop procedures for correcting many situations, resolving the counts of many GQs, and reducing the number of unresolved GQs that required count imputation or other procedures to address duplication of GQs or people living in GQs.

## Group Quarters Count Imputation

Count imputation for GQs was applied to unresolved responses in specified GQ types. As described in the Background section, a GQ was considered unresolved if we had evidence that the GQ was occupied around Census Day, but the population count for the GQ on the census processing file was 0.

We also considered imputing counts for GQs whose response count was greater than 0, but considered to be implausible. Examples include a response count of 3 or 1,500 for a dormitory whose expected and maximum counts during the Group Quarters Advance Contact (GQAC) operation were 120 and 150, respectively. However, expert review of the data after other corrective actions removed almost all of these cases from GQ count imputation. Further, we were reluctant to change a response unless we felt a preponderance of the evidence indicated it was incorrect. Thus, GQ count imputation was only applied to occupied GQs with a population count of 0. GQ count imputation was not applied to GQs with a positive count that was identified as implausible.

*Table 1* provides the types of GQs for which we did not apply count imputation for the various reasons described above.

*Table 1: Group Quarter Types in Which Count Imputation Was Not Applied*

| CODE | VALUE |
|---|---|
| 101 | Federal Detention Centers |
| 102 | Federal Prisons |
| 701 | Emergency and Transitional Shelters (With Sleeping Facilities) for People Experiencing Homelessness |
| 702 | Soup Kitchens |
| 704 | Regularly Scheduled Mobile Food Vans |
| 706 | Targeted Non-Sheltered Outdoor Locations |
| 904 | Other Noninstitutional Group Quarters |
| 999 | Unknown Group Quarters Type |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team

# Imputation Methods

For GQ count imputation we employed the following methods:

1. Ratio Imputation
2. Adjusted Residual from Facility-Level Total for College Housing
3. Percentile Imputation

## Ratio Imputation

For most unresolved GQs, the count imputation procedure relied on a ratio based on information we had already received from the GQ in the GQAC or through our current surveys, such as the American Community Survey or the Current Population Survey. *Table 2* provides auxiliary variables depicting information we used in the imputation. They are listed according to our judgment of their predictive value in the imputation.

*Table 2: Variables Used for Group Quarters Count Ratio Imputation*

| Variable Name | Description |
|---|---|
| GQAC Expected Count | The expected count of people at the Group Quarters on Census Day, as collected during GQAC |
| GQAC Max Number of People | The maximum number of people that can live or stay at the Group Quarters at a given time, as collected during GQAC |
| Current GQ Size | The number of people at the Group Quarters from the last ACS or Current Surveys visit |
| Max Number of People | The maximum number of people at the Group Quarters from ACS or Current Surveys |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team

The first two variables in *Table 2*, when available, were collected in the GQAC, conducted shortly before Census Day. The final two derive from collection in the current surveys, which would be less current, and were only available for GQs that fell into sample for one of the

current household surveys. We suspected the expected count from GQAC and the current GQ size (variables 1 and 3) would track our target—the reported count—slightly better than the maximum number of people who might live in the GQ (variables 2 and 4). Our analyses on GQs with a good response and all four variables as part of researching the 2020 GQ count imputation methodology supported our assumptions about the predictive value of the four variables.

Ratio imputation is a common method to impute counts from establishment surveys. After an intensive, but short, period of researching imputation methods for GQs, we decided to use ratio imputation as our preferred method of imputation for GQs.

Although ratio imputation is different from the hot-deck imputation used to impute housing unit counts, it was well justified for a variety of reasons. When researching imputation methods, we considered the fact that GQs vary in size considerably, the counts in just one GQ can influence block and tract census counts considerably, and population counts were available for many unresolved GQs from the GQ Advance Contact or current household surveys—such as the Current Population Survey or American Community Survey—if they fell into sample for a current household survey. Ratio imputation seemed apt at taking advantage of rich auxiliary data to impute person counts for each GQ.

When applying ratio imputation, we multiplied the auxiliary count of the unresolved GQ by a conversion factor. For each detailed GQ type within each state, we calculated a conversion factor (ratio) for each of the four main auxiliary variables:

- GQ Advance Contact (GQAC) expected count.
- GQAC maximum capacity.
- Current survey expected count.
- Current survey maximum capacity.

For example, to calculate the conversion factor for the GQAC expected count, within each state and GQ type, we started by summing the reported population counts from resolved GQs within our imputation base. We divided this sum by the total of GQAC expected counts among these GQs. We then used this ratio to convert the GQAC expected count of an unresolved GQ in this state and GQ type into a Census Day imputed count. Thus, for an unresolved College GQ with a GQAC expected count, the following equation would be used:

$$Imputed\ Population\ Count = GQAC\ Expected\ Count * \frac{\sum_{GQTYPE=College} Reported\ GQ\ Pop\ Count}{\sum_{GQTYPE=College} GQAC\ Expected\ Count}$$

We constructed ratios in the same manner using the GQAC maximum capacity, current survey expected count, and current survey maximum capacity variables. National ratios for the four variables by GQ type are shown in *Table 9*.

For the ratio imputation, we cycled through the four covariates in order until a covariate was available. The first available covariate was used to impute the population count. For example, if an unresolved GQ had a GQAC expected count, the GQAC conversion factor was used to impute the GQ population count. However, if the GQAC expected count was missing, we then checked if the GQAC maximum capacity was reported. If so, we used the GQAC maximum capacity conversion factor. Otherwise, we continued down the list until a covariate was available. If all four covariates were missing, we relied on the Adjusted Residual from Facility-Level Total imputation (for College Housing imputation only) or percentile imputation.

## Adjusted Residual from Facility-Level Total for College Housing

The second imputation approach, labeled Adjusted Residual from Facility-Level Total for College Housing imputation, was sometimes applied, but only for GQs of type 501, College and University Housing, and only when none of the four variables previously discussed was available. This hybrid method first applied ratio estimation wherever possible, and otherwise based the imputation on the Integrated Postsecondary Education Data System (IPEDS) of the National Center for Education Statistics. The IPEDS data provided total room capacity for the GQ facility—a college or university—across all its GQs locations, that is, its student housing, fraternity and sorority houses, university-owned apartments or living quarters for students, and other such residences. To use this method, we first identified all GQ locations associated with a specific university in the IPEDS. (Note that this did not include regular housing units or apartments not operated by the university in which a student might have lived.)

Within each university, we identified all its GQs that had a reported count or had an auxiliary variable that allowed us to apply the ratio estimation method previously described; at this point in the processing, those GQs were all resolved. The university's remaining GQs were still unresolved. We took the total room capacity for the university as found in the IPEDS, and then subtracted from it the population counts of the GQs that were reported or imputed. The difference, called the residual count, was then spread among the unresolved GQs in the university. Note that by accounting for populations reported or imputed before computing the residual, the count imputation operation avoided reintroducing duplicate persons already removed in prior processes.

## Percentile Imputation

A small subset of the unresolved GQs had none of the four auxiliary variables defined in *Table 2*. For GQs locations that had none of the four auxiliary variables available and were not in GQ type 501, College and University Housing, we used a percentile imputation. This method partitioned the GQ universe into imputation cells based on the detailed GQ type and U.S. state. Then, we calculated the specified percentile of the GQ population count (e.g., the 65th percentile) and imputed the unresolved GQs with that percentile of the GQ population count in

the cell. We estimated the percentile as the value that minimized the imputation bias in a cross-validation evaluation built from responding GQs.

For example, for nursing homes, we identified the 65th percentile as the value to use for imputation. Separately within each state, we computed the 65th percentile of reported counts from nursing homes. We then imputed any nursing home missing all auxiliary data with the 65th percentile from its state.

We also calculated within each GQ type the national imputation value across all states. This count was used for imputation when we did not have any eligible resolved (reported) GQs for the GQ Type in a state.

### Order of Methods

For an unresolved GQ, we always used the ratio method if one of the four covariates was available. If none of the covariates was available, and the GQ was included in the College Housing type, we next tried to use the Adjusted Residual from Facility-level Total for College Housing method. If we could not use that method, we relied on the percentile imputation. The ordering was done to favor the methods that used the most timely and accurate data about each unresolved GQ.

## Determining the Imputation Parameters

The GQ imputation procedure comprised three techniques: ratio imputation, Adjusted Residual from Facility-Level Total for College Housing imputation, and percentile imputation. For a given unresolved GQ, the technique was dependent on the presence of covariates for the unresolved GQs. This section describes how parameters for each technique were set.

### GQ Universe

The GQ universe contained 267,000 GQs; 44,500 GQs had no clear status (occupied, vacant, delete) from the fieldwork. During post-data collection processing, various activities, including the callback operation and subject-matter review, resulted in assigning a status to all the GQs. After these processing efforts, 5,500 GQs were identified as occupied with no population count.

The occupied group quarters requiring imputation included refusals without any reported people. In addition, group quarters that were open on Census Day but were vacant during the GQ Enumeration visit (which occurred between late July and mid-October 2020) required imputation.

*Table 3: GQ Universe*

| GQ Outcome | Resolved | Unresolved - Required Imputation | Total |
|---|---|---|---|
| Occupied GQ | 196,000 | 1,000 | 197,000 |
| Open on Census Day, Vacant During Visit | 19,500 | 1,900 | 21,500 |
| Refusal GQ | 5,100 | 2,600 | 7,800 |
| Vacant GQ | 30,500 | N < 15 | 30,500 |
| Delete GQ | 7,600 | N < 15 | 7,600 |
| Nonresidential | 2,500 | 0 | 2,500 |
| Total | 262,000 | 5,500 | 267,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Rows and columns may not be additive because of rounding.

*Table 3* shows counts of the GQ universe by GQ outcome and whether the GQ population had to be imputed. Determining GQ outcome involved Census Bureau staff from several operations and analyses, and was not always clear given the complexities of the GQ data collection. In general, the "occupied GQ" classification was intended to indicate that the GQ was occupied during the GQ Enumeration visit and on Census Day. The "open on Census Day, vacant during visit"[2] outcome was intended to indicate that the GQ was vacant during the GQ Enumeration visit in the summer of 2020, but occupied on Census Day. The census reference day for most GQ types was April 1, 2020, however, for colleges and universities, the reference day was before April 1, 2020, so that students would be counted at their college residence before many colleges and universities closed because of the COVID-19 pandemic. Refusals could be occupied or vacant.

Unless we had evidence to the contrary, GQs with a status of (1) occupied, (2) open on Census Day, vacant during visit, or (3) refusal were all assumed to be occupied and eligible for GQ count imputation. The remaining categories (vacant, delete, nonresidential) were not considered occupied and not eligible for GQ count imputation unless subject-matter experts determined them to be occupied during their review of the Decennial Response File 2 (refer to Thieme, 2021).

Table 4 shows the distribution of the resolved and unresolved occupied GQs by GQ type. Table 13 in the Appendix has a full list of the GQ type codes. Unlike other GQ types, shelters (GQ types 701 to 706) applied a reference date during the enumeration date in September, rather than April 1, 2020. Thus, for shelters only, the "open on Census Day, vacant during visit" outcome indicated that the shelters should have been counted as vacant. Shelters were enumerated during Service-Based Enumeration (SBE) and not GQE (refer to United States Census Bureau 2021). Given that shelters were enumerated differently than other GQs, and that a majority of the GQs without a population count were vacant during the visit, population

[2] During GQ Enumeration, the GQ was found to be vacant, but the contact at the GQ said the GQ was open on Census Day. This is different from the vacant GQs that were reported to be vacant on Census Day.

counts were not imputed for these GQs. Thus, all shelters are classified as resolved and not included in the 5,500 unresolved GQs sent to GQ count imputation.

*Table 4: Resolved and Unresolved GQ Counts by Aggregated GQ Type for Occupied GQs*

| GQ Type Group | Resolved | Unresolved - Required Imputation | Total |
|---|---|---|---|
| Unclassified | N < 15 | 0 | N < 15 |
| Correctional Facilities* | 17,000 | 200 | 17,000 |
| Juvenile Facilities | 8,700 | 400 | 9,100 |
| Nursing Facilities* | 29,500 | 350 | 30,000 |
| Hospitals | 3,000 | 200 | 3,200 |
| College Housing* | 38,000 | 1,300 | 39,500 |
| Military* | 6,000 | 100 | 6,100 |
| Shelters | 56,000 | 0 | 56,000 |
| Group Homes | 73,500 | 2,400 | 76,000 |
| Other | 29,500 | 650 | 30,000 |
| Total | 262,000 | 5,500 | 267,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Rows and columns may not be additive because of rounding.
*Denotes GQ Type was included in National Processing Center calling operation

Additionally, population counts were not imputed for the following GQ types:

- Federal Detention Centers (GQ type 101).
- Federal Prisons (GQ type 102).
- Other Noninstitutional Group Quarters (GQ type 904).
- Unknown Group Quarters (GQ type 999).

GQs with these types were all classified as resolved. For the federal detention centers and federal prisons (a subset of the Correctional Facilities GQ Type), the GQs missing a population count were determined to be vacant or included in the population count for another GQ. We established Memorandums of Understanding with the Bureau of Prisons, U.S. Immigration and Customs Enforcement (ICE), U.S. Marshal, and State prisons to provide administrative records for the 2020 Census. Often, one person reported a count for an entire facility, rather than every GQ. Thus, imputing counts would have overcounted these populations.

## Imputation Base

The imputation base is the set of resolved GQs that were eligible to be used in calculating the imputation parameters prior to performing Hidiroglou-Berthelot (HB) edits. There were 262,000 GQs with a resolved status. Of those, 81,000 were resolved with a status other than occupied or had an occupied outcome code with a population count of 0. They were not included in the imputation base.

As already mentioned, the federal detention centers, federal prisons, other noninstitutional group quarters, shelters, and unknown group quarters were not eligible for imputation. Thus, they were excluded from the imputation base as well. There were an 35,000 GQs with these GQ types.

Prior to the GQ count imputation, 100 additional GQs were given a count based on reports at the facility level. For these GQs, the facility-level counts were distributed across the GQs within the facility. GQs in these facilities were treated as resolved, but not included in the imputation base because their person counts were not directly reported.

*Table 5* shows a decomposition of the GQ universe into those that were included and excluded from the imputation base. After removing the vacant, nonexistent, and excluded GQ types, there were 145,000 GQs in the imputation base. These 145,000 GQs were then edited. The editing resulted in suppressing some GQs from being used in the calculation of some parameters.

*Table 5: Resolved GQs In and Out of the Imputation Base*

| GQ Type Group | Included in Imputation Base | Excluded from Imputation Base | Total Resolved GQs |
|---|---|---|---|
| Unclassified | 0 | N < 15 | N < 15 |
| Correctional Facilities | 12,000 | 5,100 | 17,000 |
| Juvenile Facilities | 6,000 | 2,700 | 8,700 |
| Nursing Facilities | 25,000 | 4,500 | 29,500 |
| Hospitals | 1,900 | 1,100 | 3,000 |
| College Housing | 29,500 | 8,500 | 38,000 |
| Military | 3,100 | 2,900 | 6,000 |
| Shelters | 0 | 56,000 | 56,000 |
| Group Homes | 61,500 | 12,000 | 73,500 |
| Other | 6,200 | 23,500 | 29,500 |
| Total | 145,000 | 116,000 | 262,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Rows and columns may not be additive because of rounding.

## Hidiroglou-Berthelot (HB) Edits

Some values for covariates were **suppressed** using the Hidiroglou-Berthelot (HB) edit procedure. This prevented some unusual or extreme values from adversely influencing the imputed values. The final section in the Appendix contains more information on the Hidiroglou-Berthelot (HB) edit procedure.

Only GQs in the imputation base underwent the HB edits. The HB edit procedure began by comparing reported population counts to the four auxiliary counts within GQ type to flag outliers. (Refer to the previous discussion in the section on *Ratio Imputation*.) The four auxiliary counts were:

- GQ Advance Contact (GQAC) expected count.
- GQAC maximum capacity.
- Current survey expected count.
- Current survey maximum capacity.

If the ratio between an auxiliary population count and the reported population count was determined to be an outlier by the HB edit procedure, both the reported count and auxiliary count were suppressed from calculating the imputation parameters based on the specific auxiliary variable. However, the reported count could be used in the calculation of imputation factors based on other auxiliary variables. For example, (1) the ratio of GQAC expected count to reported GQ population could be an outlier, but (2) the ratio of GQAC maximum capacity to GQ population might not be an outlier according to the HB edit procedure. In that case, the GQ would be included in the imputation base for calculating the second ratio, but not for the first ratio.

For each covariate used to calculate the imputed values, *Table 6* shows counts of the 145,000 GQs in the imputation base by type of GQ that had a missing ratio. For the ratio methods, GQs with a blank covariate were not used in calculating the imputation parameter for the ratio based on that covariate. To be eligible for the percentile method imputation base, all four covariates had to be filled.

*Table 6: Counts of GQs with Missing Values by GQ Type*

| GQ Type Group | Total Resolved GQs with Missing Values | | | | |
|---|---|---|---|---|---|
| | GQAC Expected Count | GQAC Maximum Capacity | Current GQ Expected Count | Current Survey Maximum Capacity | Percentile Method |
| Correctional Facilities | 8,600 | 4,400 | 4,700 | 3,600 | 8,900 |
| Juvenile Facilities | 2,200 | 1,500 | 2,400 | 350 | 3,100 |
| Nursing Facilities | 5,900 | 4,200 | 9,100 | 400 | 10,500 |
| Hospitals | 750 | 550 | 1,000 | 60 | 1,100 |
| College Housing | 11,500 | 7,700 | 12,500 | 1,200 | 15,500 |
| Military | 2,100 | 1,500 | 1,600 | 400 | 2,200 |
| Group Homes | 28,500 | 23,500 | 32,500 | 2,200 | 36,500 |
| Other | 3,000 | 2,400 | 3,000 | 200 | 3,800 |
| Total | 62,500 | 45,500 | 67,000 | 8,400 | 82,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Columns may not be additive because of rounding.

*Table 7* shows counts of GQs in the imputation base that have a non-missing covariate value, but failed an HB edit (refer to the final section in the Appendix for details on how the HB edit procedure identifies outliers). This table shows the number of GQs with a value that was suppressed and not used in calculating the imputation ratio. For the percentile method, GQs with at least one suppressed covariate were excluded from calculating the percentile imputation parameter.

*Table 7: Counts of GQs with Suppressed Values by GQ Type*

| GQ Type Group | Total Resolved GQs that Failed HB Edits | | | | |
|---|---|---|---|---|---|
| | GQAC Expected Count | GQAC Maximum Capacity | Current Survey Expected Count | Current Survey Maximum Capacity | Percentile Method |
| Correctional Facilities | 50 | 100 | 250 | 250 | 150 |
| Juvenile Facilities | 60 | 80 | 60 | 100 | 100 |
| Nursing Facilities | 200 | 150 | 250 | 400 | 250 |
| Hospitals | 30 | 50 | 40 | 70 | 50 |
| College Housing | 900 | 550 | 400 | 850 | 700 |
| Military | 30 | 60 | 60 | 60 | 60 |
| Group Homes | 450 | 200 | 350 | 300 | 500 |
| Other | 30 | 80 | 100 | 150 | 150 |
| Total | 1,700 | 1,300 | 1,600 | 2,200 | 1,900 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Columns may not be additive because of rounding.

*Table 8* shows counts of GQs by type that contributed to calculating the numerator and denominator for each imputation parameter. For example, 81,000 GQs contributed to calculating the GQAC expected count ratio.

*Table 8: Counts of GQs Used to Create Ratio Factors by GQ Type*

| GQ Type Group | Total Resolved GQs Used in Each Calculation | | | | |
|---|---|---|---|---|---|
| | GQAC Expected Count | GQAC Maximum Capacity | Current Survey Expected Count | Current Survey Maximum Capacity | Percentile Method |
| Correctional Facilities | 3,300 | 7,500 | 7,000 | 8,100 | 2,900 |
| Juvenile Facilities | 3,700 | 4,400 | 3,500 | 5,600 | 2,900 |
| Nursing Facilities | 19,000 | 20,500 | 15,500 | 24,000 | 14,500 |
| Hospitals | 1,200 | 1,300 | 900 | 1,800 | 750 |
| College Housing | 17,000 | 21,500 | 16,500 | 27,500 | 13,000 |
| Military | 950 | 1,500 | 1,500 | 2,600 | 800 |
| Group Homes | 32,500 | 38,000 | 28,500 | 59,000 | 24,500 |
| Other | 3,200 | 3,800 | 3,100 | 5,800 | 2,300 |
| Total | 81,000 | 98,500 | 77,000 | 135,000 | 61,500 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Columns may not be additive because of rounding.

Note, the HB edit took the GQ size into account when identifying outliers—the ratios were transformed so that more importance was placed on a small deviation from the median ratio for a large GQ as opposed to a large deviation for a small GQ (Hidiroglou and Berthelot, 1986).

## Calculation of Parameters

Global parameters were calculated for the ratio and percentile techniques. *For the ratio technique*, ratios were calculated at the state and national levels. First, a state-level ratio was calculated by GQ type for each of the four variables. To use this ratio, we had to know the GQ type and state of the unresolved GQ. Second, a national ratio was calculated by GQ type for each variable. Within GQ type, values were summed across all states. To use this value, it was only necessary to know the GQ type of the unresolved GQ. *Table 9* shows the national imputation parameters for each GQ type.

*For the percentile technique*, GQs in the imputation base were sorted by GQ count, and a predetermined percentile of this distribution was used as the imputation parameter (value). This was done by state and nationally. The full classification of GQ type was used in calculating ratios and percentiles, while collapsed groupings were used for the HB edits.

*Table 9: National GQ Imputation Ratio Parameters (all values are multiplied by 100 in this table)*

| GQ Type | Ratio of GQAC Expected Count | Ratio of GQAC Maximum Capacity | Ratio of Current Survey Expected Count | Ratio of Current Survey Maximum Capacity |
|---|---|---|---|---|
| State Prisons | 83 | 72 | 89 | 72 |
| Local Jails and Other Municipal Confinement Facilities | 71 | 57 | 80 | 58 |
| Correctional Residential Facilities | 72 | 60 | 84 | 62 |
| Military Disciplinary Barracks and Jails | N < 15 | N < 15 | N < 15 | 60 |
| Group Homes for Juveniles (noncorrectional) | 75 | 62 | 83 | 63 |
| Residential Treatment Centers for Juveniles (noncorrectional) | 78 | 67 | 83 | 67 |
| Correctional Facilities Intended for Juveniles | 60 | 46 | 67 | 45 |
| Nursing Facilities/Skilled Nursing Facilities | 87 | 74 | 93 | 74 |
| Mental (Psychiatric) Hospitals and Psychiatric Units in Other Hospitals | 86 | 78 | 92 | 76 |
| Hospitals with Patients Who Have No Usual Home Elsewhere | 77 | 65 | 100 | 64 |
| In-Patient Hospice Facilities | 79 | 67 | 82 | 67 |
| Military Treatment Facilities with Assigned Patients | N < 15 | N < 15 | N < 15 | N < 15 |
| Residential Schools for People with Disabilities | 72 | 60 | 84 | 61 |
| College/University Student Housing (College/University owned/leased/managed) | 91 | 80 | 94 | 81 |

| GQ Type | Ratio of GQAC Expected Count | Ratio of GQAC Maximum Capacity | Ratio of Current Survey Expected Count | Ratio of Current Survey Maximum Capacity |
|---|---|---|---|---|
| Military Quarters | 77 | 70 | 99 | 63 |
| Military Ships | N/A | N/A | 51 | 30 |
| Group Homes Intended for Adults | 90 | 82 | 95 | 81 |
| Residential Treatment Centers for Adults | 78 | 68 | 81 | 70 |
| Maritime/Merchant Vessels | N/A | N/A | 46 | 27 |
| Workers' Group Living Quarters and Job Corps Centers | 77 | 58 | 80 | 62 |
| Living Quarters for Victims of Natural Disaster | N < 15 | 70 | N < 15 | 60 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team

## Capping Auxiliary Variables for Unresolved GQs

The auxiliary variables for the unresolved GQs were reviewed for extreme values. First, we calculated the 10th (and 90th) percentiles of the four auxiliary variables by GQ type and state. If an auxiliary variable for the GQ was below the 10th percentile (or above the 90th percentile), we capped the value at the 10th (or 90th) percentile. The values of 10th (and 90th) percentiles were selected after reviewing the distribution of the auxiliary variables. This capping prevented us from imputing extremely small or large GQ populations. However, it may have added some bias for very large or small GQs. For example, we might have capped a reported GQAC expected count of 10,000 to be 1,000. Then, the capped 1,000 GQAC expected count was used to determine the imputed value. Some of the very large auxiliary reported values for the unresolved GQs were erroneous. Imputing a GQ population directly from the reported value would have given a large amount of influence to the GQ. Given the limited time to investigate these atypical GQs, capping the auxiliary values provided a safeguard against imputing an extreme value without additional research into the GQ.

# Results

*Table 10* shows that most GQs were imputed using one of the four ratios. In fact, 100 GQs were imputed using IPEDS and 200 GQs were imputed using the percentile method. Together these two methods make up about 5 percent of the imputed GQs. Because most of the unresolved GQs had a state code, most GQs used state-specific parameters for their GQ type.

*Table 10: Imputation Method*

| GQ Imputation Method | Unresolved GQs | GQ People Imputed |
|---|---|---|
| GQAC Expected Count (GQ Type by State) | 3,000 | 106,000 |
| GQAC Expected Count (GQ Type) | N < 15 | 80 |
| GQAC Maximum Capacity (GQ Type by State) | 650 | 19,000 |
| GQAC Maximum Capacity (GQ Type) | N < 15 | 150 |
| Current Survey Expected Count (GQ Type by State) | 20 | 800 |
| Current Survey Expected Count (GQ Type) | 0 | N < 15 |
| Current Survey Maximum Capacity (GQ Type by State) | 1,500 | 31,500 |
| Current Survey Maximum Capacity (GQ Type) | N < 15 | 20 |
| Adjusted Residual from Facility-Level Total for College Housing | 100 | 8,600 |
| Percentile Method (GQ Type by State) | 200 | 2,500 |
| Percentile Method (GQ Type) | N < 15 | 550 |
| Total | 5,500 | 169,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Columns may not be additive because of rounding. Note that noise was injected into the GQ People imputed column for disclosure avoidance, whether there were any unresolved GQs in the row or not.

*Table 11* shows how many GQs and people were imputed by GQ type. As a result of the GQ imputation effort, a population count was imputed for 5,500 GQs. The total of 169,000 people were added to those 5,500 GQs. This increased the GQ population from 8,191,000 to 8,360,000. From *Table 11*, we see that the average number of people per GQ varies considerably by the GQ Type, providing evidence for creating specific imputation parameters for each GQ type. Additionally, we see the number of GQs needing imputation varied across the GQ types. Because the ratio imputation relied on previously reported GQ-specific data, the average number of people per GQ may differ between the resolved and imputed GQs.

*Table 11: Summary of Imputations by GQ Type*

| GQ Type | GQs | | | GQ People | | |
|---|---|---|---|---|---|---|
| | Resolved | Imputed | Total | Resolved | Imputed | Total |
| State Prisons | 9,800 | 60 | 9,800 | 1,152,000 | 7,300 | 1,159,000 |
| Local Jails and Other Municipal Confinement Facilities | 3,500 | 70 | 3,600 | 491,000 | 8,700 | 499,000 |
| Correctional Residential Facilities | 1,200 | 50 | 1,200 | 65,500 | 2,700 | 68,500 |
| Military Disciplinary Barracks and Jails | 30 | N < 15 | 40 | 1,500 | 250 | 1,700 |
| Group Homes for Juveniles (noncorrectional) | 4,600 | 250 | 4,800 | 33,000 | 1,900 | 35,000 |
| Residential Treatment Centers for Juveniles (noncorrectional) | 2,400 | 100 | 2,500 | 27,500 | 1,700 | 29,000 |
| Correctional Facilities Intended for Juveniles | 1,800 | 70 | 1,800 | 26,500 | 1,200 | 27,500 |
| Nursing Facilities/Skilled Nursing Facilities | 29,500 | 350 | 30,000 | 1,613,000 | 20,000 | 1,633,000 |
| Mental (Psychiatric) Hospitals and Psychiatric Units in Other Hospitals | 1,100 | 70 | 1,200 | 39,500 | 3,400 | 43,000 |

| GQ Type | GQs | | | GQ People | | |
|---|---|---|---|---|---|---|
| | Resolved | Imputed | Total | Resolved | Imputed | Total |
| Hospitals with Patients Who Have No Usual Home Elsewhere | 450 | 40 | 450 | 9,400 | 1,200 | 10,500 |
| In-Patient Hospice Facilities | 750 | 50 | 800 | 9,000 | 700 | 9,700 |
| Military Treatment Facilities with Assigned Patients | 30 | N < 15 | 40 | 1,100 | 400 | 1,500 |
| Residential Schools for People with Disabilities | 700 | 40 | 750 | 9,200 | 550 | 9,700 |
| College/University Student Housing (College/University owned/leased/managed) | 38,000 | 1,300 | 39,500 | 2,712,000 | 81,000 | 2,793,000 |
| Military Quarters | 5,700 | 100 | 5,800 | 328,000 | 8,300 | 337,000 |
| Military Ships | 250 | 0 | 250 | 32,500 | 0 | 32,500 |
| Group Homes Intended for Adults | 62,500 | 1,900 | 64,500 | 521,000 | 14,000 | 536,000 |
| Residential Treatment Centers for Adults | 11,000 | 500 | 11,500 | 154,000 | 8,700 | 163,000 |
| Maritime/Merchant Vessels | 350 | 0 | 350 | 2,400 | 0 | 2,300 |
| Workers' Group Living Quarters and Job Corps Centers | 11,500 | 650 | 12,000 | 83,000 | 7,500 | 90,500 |
| Living Quarters for Victims of Natural Disaster | 80 | N < 15 | 80 | 250 | N < 15 | 250 |
| Not Eligible for GQ Count Imputation | 76,000 | 0 | 76,000 | 879,000 | 0 | 879,000 |
| Total | 262,000 | 5,500 | 267,000 | 8,190,000 | 170,000 | 8,360,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Rows and columns may not be additive because of rounding.

*Table 12* shows how many GQs and people were imputed in each state. Overall, about 2 percent of the GQs were imputed with a GQ count. About 2 percent of the GQ population was imputed. The states varied in their imputation rates and the number of people imputed. In general, the imputation parameters were calculated separately for each state and GQ type because the average number of people per GQ varied across states and GQ types. The GQ count imputation had a large impact on the total GQ person count in Delaware. A small number of unresolved GQs with a large population could lead to significant undercoverage of the GQ population, if not imputed. Although very important, the GQ population in each state tends to be much smaller than the housing unit population.

*Table 12: Summary of Imputations by State*

| State | GQs | | | GQ People | | |
|---|---|---|---|---|---|---|
| | Resolved | Imputed | Total | Resolved | Imputed | Total |
| AL | 3,000 | 150 | 3,100 | 122,000 | 6,400 | 128,000 |
| AK | 2,800 | 100 | 2,900 | 30,000 | 1,200 | 31,500 |
| AZ | 4,800 | 200 | 5,000 | 156,000 | 5,100 | 161,000 |
| AR | 2,000 | 30 | 2,100 | 80,000 | 2,900 | 83,000 |
| CA | 40,500 | 550 | 41,000 | 940,000 | 18,500 | 959,000 |
| CO | 3,300 | 70 | 3,400 | 126,000 | 1,300 | 127,000 |
| CT | 4,100 | 20 | 4,100 | 108,000 | 550 | 109,000 |

| State | GQs | | | GQ People | | |
|---|---|---|---|---|---|---|
| | Resolved | Imputed | Total | Resolved | Imputed | Total |
| DE | 550 | 20 | 600 | 20,000 | 2,600 | 23,000 |
| DC | 850 | N < 15 | 850 | 40,500 | 90 | 41,000 |
| FL | 11,000 | 450 | 11,500 | 450,000 | 16,500 | 467,000 |
| GA | 4,900 | 100 | 5,000 | 251,000 | 5,200 | 256,000 |
| HI | 2,100 | N < 15 | 2,100 | 41,000 | 200 | 41,500 |
| ID | 1,500 | 20 | 1,500 | 49,500 | 250 | 50,000 |
| IL | 6,800 | 150 | 6,900 | 274,000 | 2,800 | 277,000 |
| IN | 4,300 | 30 | 4,400 | 178,000 | 1,200 | 179,000 |
| IA | 3,500 | 100 | 3,600 | 96,500 | 2,200 | 99,000 |
| KS | 2,500 | 20 | 2,500 | 86,500 | 500 | 87,000 |
| KY | 2,600 | 60 | 2,700 | 123,000 | 2,700 | 126,000 |
| LA | 2,500 | 100 | 2,600 | 119,000 | 5,400 | 125,000 |
| ME | 2,100 | 100 | 2,200 | 35,000 | 1,900 | 36,500 |
| MD | 4,800 | 70 | 4,800 | 125,000 | 1,500 | 127,000 |
| MA | 7,800 | 90 | 7,800 | 246,000 | 1,400 | 247,000 |
| MI | 8,500 | 450 | 9,000 | 212,000 | 10,500 | 223,000 |
| MN | 7,500 | 30 | 7,500 | 139,000 | 450 | 139,000 |
| MS | 1,500 | 90 | 1,600 | 87,500 | 6,400 | 94,000 |
| MO | 5,600 | 200 | 5,800 | 163,000 | 6,700 | 170,000 |
| MT | 1,300 | 30 | 1,300 | 29,000 | 550 | 29,500 |
| NE | 1,800 | N < 15 | 1,800 | 49,500 | 150 | 49,500 |
| NV | 1,200 | 20 | 1,200 | 36,500 | 650 | 37,000 |
| NH | 1,500 | N < 15 | 1,600 | 43,000 | 80 | 43,000 |
| NJ | 5,900 | 40 | 5,900 | 180,000 | 1,200 | 181,000 |
| NM | 1,900 | 50 | 2,000 | 42,000 | 1,400 | 43,000 |
| NY | 22,000 | 250 | 22,000 | 608,000 | 5,900 | 614,000 |
| NC | 8,300 | 400 | 8,700 | 278,000 | 8,800 | 287,000 |
| ND | 850 | N < 15 | 850 | 26,500 | 80 | 26,500 |
| OH | 7,700 | 100 | 7,800 | 298,000 | 3,400 | 301,000 |
| OK | 2,600 | N < 15 | 2,600 | 117,000 | 650 | 118,000 |
| OR | 5,600 | 80 | 5,700 | 96,500 | 1,600 | 98,000 |
| PA | 11,500 | 200 | 11,500 | 400,000 | 6,500 | 406,000 |
| RI | 1,200 | N < 15 | 1,200 | 45,500 | 80 | 46,000 |
| SC | 3,400 | 150 | 3,500 | 134,000 | 4,800 | 139,000 |
| SD | 800 | N < 15 | 800 | 32,000 | 300 | 32,000 |
| TN | 4,000 | 150 | 4,100 | 152,000 | 5,700 | 157,000 |
| TX | 11,000 | 200 | 11,500 | 596,000 | 13,000 | 609,000 |
| UT | 2,000 | 30 | 2,000 | 56,500 | 300 | 56,500 |
| VT | 900 | N < 15 | 950 | 25,000 | 250 | 25,000 |
| VA | 5,600 | 100 | 5,700 | 239,000 | 1,600 | 240,000 |
| WA | 7,500 | 200 | 7,700 | 159,000 | 4,300 | 163,000 |

| State | GQs | | | GQ People | | |
|---|---|---|---|---|---|---|
| | Resolved | Imputed | Total | Resolved | Imputed | Total |
| WV | 1,300 | 20 | 1,300 | 52,000 | 250 | 52,500 |
| WI | 6,400 | 100 | 6,600 | 150,000 | 1,900 | 152,000 |
| WY | 750 | N < 15 | 800 | 13,000 | 250 | 13,500 |
| PR | 3,100 | 40 | 3,200 | 36,000 | 1,300 | 37,500 |
| Total | 262,000 | 5,500 | 267,000 | 8,190,000 | 169,000 | 8,360,000 |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team
Note: Rows and columns may not be additive because of rounding.

# Conclusions

Imputing a population size for occupied GQs without a response required numerous steps. First, resolved GQs were edited to remove outliers and influential observations from the base used to calculate imputation parameters. The HB edits were applied separately for the four main covariates: GQAC expected count, GQAC maximum capacity, Current Survey expected count, and Current Survey maximum capacity. After carefully editing the imputation base, state-specific and national imputation parameters were calculated by GQ type. When present, auxiliary variables for unresolved GQs were top- and bottom-coded to prevent extremely small or large imputations. After this editing of the auxiliary variables for unresolved GQs, the population size was imputed.

The population size for 5,500 GQs was imputed. This is about 2 percent of the GQ universe, although the unresolved rate varied by GQ type and state. Imputations resulted in adding 169,000 people to the 5,500 GQs.

# References

Hidiroglou, M.A., and Berthelot, J.M. (1986), "Statistical Editing and Imputation for Periodic Business Surveys," Survey Methodology, 12, 73-83.

Thieme, M. (2021), "Finding 'Anomalies' Illustrates 2020 Census Quality Checks Are Working," Random Samplings Blog, March 09, 2021. https://www.census.gov/newsroom/blogs/random-samplings/2021/03/finding_anomalies.html (accessed January 2023).

United States Census Bureau (2021), "Group Quarters Operation Detailed Operational Plan v2.0," September 9, 2021. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/GQ_detailed_operational_plan-v2.pdf (accessed January 2023).

United States Census Bureau (2022a), "Updates to the 2020 Census Response Processing Operation Detailed Operational Plan," January 6, 2022. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/RPO_detailed_operational_plan-v2.pdf (accessed January 2023).

United States Census Bureau (2022b), “Census 2020: High-Level Overview of Group Quarters Count Imputation,” March 3, 2022. https://www2.census.gov/programs-surveys/decennial/2020/program-management/memo-series/2020-memo-2022_08.pdf (accessed January 2023).

# Appendix

## GQ Type Codes

*Table 13: Group Quarter Types*

| CODE | GQ Type | GQ Group |
|---|---|---|
| 000 | Unassigned | Unclassified |
| 010 | Campground | Unclassified |
| 020 | Recreational Vehicle (RV) Park | Unclassified |
| 030 | Marina | Unclassified |
| 040 | Hotel or Motel | Unclassified |
| 050 | Racetrack | Unclassified |
| 060 | Circus or Carnival | Unclassified |
| 090 | Other Transitory Location | Unclassified |
| 101 | Federal Detention Centers | Correctional Facilities |
| 102 | Federal Prisons | Correctional Facilities |
| 103 | State Prisons | Correctional Facilities |
| 104 | Local Jails and Other Municipal Confinement Facilities | Correctional Facilities |
| 105 | Correctional Residential Facilities | Correctional Facilities |
| 106 | Military Disciplinary Barracks and Jails | Correctional Facilities |
| 201 | Group Homes for Juveniles (noncorrectional) | Juvenile Facilities |
| 202 | Residential Treatment Centers for Juveniles (noncorrectional) | Juvenile Facilities |
| 203 | Correctional Facilities Intended for Juveniles | Juvenile Facilities |
| 301 | Nursing Facilities/Skilled Nursing Facilities | Nursing Facilities |
| 401 | Mental (Psychiatric) Hospitals and Psychiatric Units in Other Hospitals | Hospitals |
| 402 | Hospitals with Patients Who Have No Usual Home Elsewhere | Hospitals |
| 403 | In-Patient Hospice Facilities | Hospitals |
| 404 | Military Treatment Facilities with Assigned Patients | Hospitals |
| 405 | Residential Schools for People with Disabilities | Hospitals |
| 501 | College/University Student Housing (College/University owned/leased/managed) | College Housing |
| 502 | College/University Student Housing (Privately owned/leased/managed) | College Housing |
| 601 | Military Quarters | Military |
| 602 | Military Ships | Military |
| 701 | Emergency and Transitional Shelters (With Sleeping Facilities) for People Experiencing Homelessness | Shelters |
| 702 | Soup Kitchens | Shelters |
| 704 | Regularly Scheduled Mobile Food Vans | Shelters |
| 706 | Targeted Non-Sheltered Outdoor Locations | Shelters |
| 801 | Group Homes Intended for Adults | Group Homes |
| 802 | Residential Treatment Centers for Adults | Group Homes |

| CODE | GQ Type | GQ Group |
|---|---|---|
| 900 | Maritime/Merchant Vessels | Other |
| 901 | Workers' Group Living Quarters and Job Corps Centers | Other |
| 903 | Living Quarters for Victims of Natural Disaster | Other |
| 904 | Other Noninstitutional Group Quarters | Other |
| 999 | Unknown Group Quarters Type | Other |

Source: U.S. Census Bureau, 2020 Census, GQ Count Imputation Team

## Hidiroglou-Berthelot (HB) Edits

The Hidiroglou-Berthelot (HB) edit detects outliers based on the ratio of two variables. In calculating the HB statistic, the ratio is transformed once to ensure that outliers are identified at both tails of the HB statistic's distribution, then transformed again to account for the size of the reporting unit. This results in identifying the records whose data exhibit the most unusual differences between the numerator and denominator variables as well as those that have more impact on the totals. These data are identified as requiring analyst review, suppression from the imputation base, or imputation.

For our purposes in this project, the HB statistic was calculated as follows.

First, we calculated the ratio between the reported population count and the auxiliary population count for each GQ *i* with positive counts for both values.

$$R_i = {}^{x_i}/_{y_i}$$

$$x_i = Reported\ Population\ Count$$

$$y_i = Auxiliary\ Population\ Count$$

We then transformed the ratios in order to detect outliers at both tails of the distribution. We calculated median ratios within each GQ type.

$$S_i = \begin{cases} 1 - \dfrac{R_{med}}{R_i} & 0 < R_i < R_{med} \\ \dfrac{R_i}{R_{med}} - 1 & R_i > R_{med} \end{cases}$$

$$R_{med} = median\ R_i$$

We then scaled the transformed ratios by GQ size to calculate the HB statistic.

$$E_i = S_i * \sqrt{\{\max(x_i, y_i)\}}$$

To detect outliers, we calculated the following values.

$$D_{Q1} = max\{E_{med} - E_{Q1}, |.05 * E_{med}|\}$$

$$D_{Q3} = max\{E_{Q3} - E_{med}, |.05 * E_{med}|\}$$

$$E_{med} = the\ median\ value\ of\ the\ HB\ statistic\ within\ GQ\ type$$

$$E_{Q1} = the\ first\ quartile\ of\ the\ HB\ statistic\ within\ GQ\ type$$

$$E_{Q3} = the\ third\ quartile\ of\ the\ HB\ statistic\ within\ GQ\ type$$

The outliers fall outside the following range.

$$\{E_{med} - c_j * D_{Q1}, E_{med} + c_j * D_{Q3}\}$$

$$c_j = parameter\ that\ controls\ the\ width\ of\ the\ acceptance\ interval$$

We set three parameter values for each GQ type: $c_1$, $c_2$, and $c_3$. These three values determined the bounds for the review, suppress, and impute flags. We conducted a manual review by plotting $x_i$ and $y_i$ values to set the bounds by GQ type. Note, this review is somewhat subjective, but imitates common practice for establishment surveys.